Appendix E

Model-based (Synthetic) Estimation

U.S. Department of Health & Human Services
Public Health Service
HRSA
Health Resources & Services Administration
Maternal & Child Health Bureau

MCHB
Maternal and Child Health Bureau

Standardization of Rates and Synthetic Estimation

Deborah Rosenberg, MS, MPH
University of Illinois School of Public Health

Originally prepared as part of the
Enhanced Analytic Skills Training Program (EASP) Training Modules
DEAL MCH Grant # MCU-17-D501-03-1

❑ Confounding and Standardization of Rates

In analytic epidemiology, when the purpose is to investigate a potentially etiologic, or causal, association between a risk factor and an outcome, "controlling for" confounding is an essential step in obtaining an unbiased estimate of the strength of the hypothesized relationship. In a public health context, when the purpose is to compare health status in different populations—when membership in a population (such as residence in a state) is the "exposure"—standardization of rates is often carried out in an effort to insure that the comparison is made accounting for differences in fundamental, structural characteristics. In either case, the goal is to account for any mixing of a third factor (or multiple other factors) with the primary association of interest.

Several terms, then, are used to describe methods that address potential confounding. Each term is typically applied in a particular analytic context, but the goal of accounting for confounding factors is the same.

"Standardization"

"Adjustment"

"Controlling for"

"Stratified Analysis"

The process of accounting for confounding involves separating the data into a series of strata and then applying a method that yields a summary or average estimate, weighted by the distribution of observations across the strata.

Standardization of rates has been most commonly used in epidemiology when comparing the mortality experience across populations. Typically, the comparisons are between large populations, often between nations, or sometimes between states. The objective is to adjust for societal level features that fundamentally distinguish the populations being compared. The age structure of a population, for instance, is a marker of social, economic, and political development in the broadest sense. Following is a series of age categories that might be used to produce an age-standardized mortality rate:

< 1

1-4

5-14

15-24

25-44

45-64

65-84

> 85

Since chronological age is positively correlated with most chronic diseases and since populations often have differing age structures, age meets the definition of a confounder and age standardized (or age-adjusted) rates are generated in order to make comparisons. For example, rates of coronary heart disease are known to increase with age. If a comparison is to be made between two geographic areas, one with 12% of its population over the age of 65, the other with only 6% of its population over the age of 65, higher death rates from CHD would be expected in the first population on the basis of this age difference alone. From a public health perspective, it is much more relevant to know if there is a difference in death rates in the two areas due to factors beyond the aging process itself--factors that may be amenable to public health interventions.

What typically distinguishes standardization of rates from other stratified methods for control of confounding is its use of an external standard. In other words, in addition to having data on the populations of interest, data from another population is used as a common benchmark. For example, the World Population, or the U.S. population from the 1940 Census, or from the 1970 Census might be used to compare U.S. and Canadian mortality rates in 1990. Sometimes an external standard may include the populations being compared, as when state rates are standardized according to national data.

In order to illustrate the process of rate standardization with an external standard, let's consider the following example comparing neonatal mortality rates in two hypothetical hospital groups in a region: one group is comprised of tertiary care hospitals with neonatal intensive care units (NICUs), the other is comprised of community hospitals which appropriately transfer the majority of high risk pregnant women for delivery at a tertiary center. The birthweight distribution in the two groups is considered a potential confounder since the tertiary care hospitals by definition serve higher risk pregnant women than do the community hospitals and birthweight is also known to be the major predictor of neonatal mortality. All live births in the state will be used as the external standard.

The data are as follows:

1. Hospital Group A: Community Hospitals

Birthweight Strata	Deaths	Births	% of Total	Stratum Specific Rate	Crude Rate
				per 1000	per 1000
< 1500	53	150	1	353.3	
1500 – 2499	12	750	5	16.0	
>= 2500	28	14,100	94	2.0	
	93	15,000	100		6.2

2. Hospital Group B: Hospitals with NICUs

Birthweight Strata	Deaths	Births	% of Total	Stratum Specific Rate	Crude Rate
				per 1000	per 1000
< 1500	65	300	2	216.7	
1500 – 2499	14	1,200	8	11.7	
>= 2500	14	13,500	90	1.0	
	93	15,000	100		6.2

External Standard: All Live Births in the State

Birthweight Strata	Deaths	Births	% of Total	Stratum Specific Rate	Crude Rate
				per 1000	per 1000
< 1500	1,375	5,000	1	275.0	
1500 – 2499	490	35,000	7	14.0	
>= 2500	460	460,000	92	1.0	
	2,325	500,000	100		4.65

The crude relative risk of neonatal death (not accounting for the birthweight distribution) when the two hospital groups are compared is 1 (6.2/6.2). Standardizing by birthweight will help determine if this relationship is a fair reflection of the neonatal mortality experience in the two hospital groups.

Standardization can be accomplished in two different ways:

> *Direct* standardization applies the stratum specific rates of each population to the number of individuals in the corresponding stratum in the standard population. This method yields an adjusted relative risk. The method is called "direct" because it uses the actual morbidity or mortality rates of the populations being compared.
>
> *Indirect* standardization, on the other hand, applies the stratum specific rates of the standard population to the number of individuals in the corresponding stratum in each of the populations being compared. This method is called "indirect" because nowhere does it use the actual morbidity or mortality rates of the populations being compared. Instead of an adjusted relative risk, indirect standardization yields standardized morbidity or mortality ratios (SMRs), one for each population being compared. Direct estimates are preferable to indirect ones, but the indirect method is used when the rates from the populations being compared are based on small numbers and therefore considered unreliable.

Direct Standardization

Using the numbers from the entire state and the rates for the two hospital groups shown in the hypothetical data above, a directly standardized relative risk is calculated as follows:

$$\textbf{Adjusted Rate for the Community Hospitals :}$$

$$= \frac{\mathbf{5{,}000 \times 353.3 + 35{,}000 \times 16.0 + 460{,}000 \times 2.0}}{\mathbf{500{,}000}}$$

$$\mathbf{= 6.5}$$

$$\textbf{Adjusted Rate for Tertiary Care Hospitals :}$$

$$= \frac{\mathbf{5{,}000 \times 216.7 + 35{,}000 \times 11.7 + 460{,}000 \times 1.0}}{\mathbf{500{,}000}}$$

$$\mathbf{= 3.9}$$

$$\textbf{Crude Relative Risk} = \frac{\textbf{6.2}}{\textbf{6.2}} = \textbf{1.0}$$

$$\textbf{Standardized Relative Risk} = \frac{\textbf{6.5}}{\textbf{3.9}} = \textbf{1.7}$$

In the process of calculating an adjusted relative risk, you can see that it is necessary to calculate what appear to be adjusted rates for each population. These recalculated rates, however, reflect what would be expected in the standard population if it had the morbidity or mortality experience of the populations being compared. In this example, the adjusted rates of 6.5 and 3.9, therefore, do not reflect the real mortality risk in the two hospital groups. They are byproducts of the standardization procedure and should not be used as stand-alone measures. In practice, these rates are sometimes reported despite the fact that they can lead to misleading and inappropriate conclusions.

The adjusted relative risk of 1.7 shows that the community hospitals have elevated neonatal mortality compared to the tertiary care centers even though the unadjusted relative risk was 1. Without adjustment, the better survival of neonates born in the tertiary care centers was masked due to the disparity in the birthweight distribution of the infants served by the two hospital groups. While the tertiary care centers have a higher incidence of low birthweight births than the community hospitals (10% v. 6 %), they have a lower neonatal mortality rate within each birthweight stratum (216.7 v. 353.0, 11.7 v.16.0, and 1.0 v. 2.0).

Indirect Standardization

Using the rates for the entire state and the numbers for the two hospital groups shown in the hypothetical data above, two standardized mortality ratios are calculated as follows:

SMR for the Community Hospitals

$$= \frac{\frac{\textbf{93}}{\textbf{15,000}} \times \textbf{1,000}}{\frac{\textbf{150} \times \textbf{275} + \textbf{750} \times \textbf{14.0} + \textbf{14,100} \times \textbf{1.0}}{\textbf{15,000}}}$$

$$= \frac{\textbf{6.2}}{\textbf{4.4}}$$

$$= \textbf{1.4}$$

SMR for the Tertiary Care Centers

$$\frac{\dfrac{93}{15{,}000}\times 1{,}000}{\dfrac{300\times 275+1{,}200\times 14.0+13{,}500\times 1.0}{15{,}000}}$$

$$=\frac{6.2}{7.5}$$

$$\mathbf{= 0.8}$$

With indirect standardization, each SMR is itself an adjusted relative risk; the numerator is the observed crude rate in a population and the denominator is its expected rate given the neonatal mortality experience in the standard population. An SMR > 1 indicates higher rates than expected and an SMR < 1 indicates lower rates than expected. The two SMRs in this example lead to the same conclusion as did the adjusted relative risk—after accounting for birthweight the community hospitals have higher neonatal mortality than do the tertiary care centers.

It is not technically correct to create a ratio of two SMRs and interpret it as a relative risk; each SMR is itself a relative risk, and as such its value is compared to a value of 1. In practice, however, SMRs are sometimes compared in this fashion. In this example, the ratio of the two SMRs is 1.4/0.8 or 1.75, very similar to the value of 1.7 for the adjusted relative risk obtained from direct standardization.

Now that the process of direct and indirect standardization has been illustrated using a particular external standard, it is important to recognize that equivalent results would be obtained with some other external standard or even with an internal standard. For example, a national hospital dataset could have been used as an external standard, or one of the two hospital groups or the two hospital groups combined could have been used as an internal standard. While the choice of different standards will impact intermediate calculations, the adjusted relative risks or SMRs will all lead to the same interpretation. In fact, the adjusted relative risks from direct standardization will be the same regardless of the standard used.

The advantage of using a common external standard, either with direct or indirect standardization, is that many geographic areas and many time periods can be compared. For example, if the neonatal mortality rates of hospital groups in another region had been compared by standardizing for birthweight with the same statewide data as was used above, it would also be possible to compare the community hospitals across regions or the tertiary care hospitals across regions. If internal standards were used, the results for the two regions would not be immediately comparable.

Interestingly, stratified analysis as typically conducted in epidemiologic studies when controlling for confounding and assessing effect modification is equivalent to direct standardization with all of the observed data combined serving as the internal standard. For instance, following are the same hospital group data organized into a set of 2x2 tables for stratified analysis. There is one table for each birthweight stratum, the "exposure" is delivery in a community hospital v. delivery in a tertiary care center, and the outcome is neonatal death v. neonatal survivor.

Stratum 1
Birthweight < 1,500 grams

		Neonatal Death Y	Neonatal Death N	
HOSP	A	53	97	150
	B	65	235	300
		118	332	450

Stratum 2
Birthweight 1,500-2,499 grams

		Neonatal Death Y	Neonatal Death N	
HOSP	A	12	738	750
	B	14	1,186	1,200
		26	1,924	1,950

Stratum 3
Birthweight >= 2,500 grams

Neonatal Death
Y N

HOSP	A	28	14,072	14,100
	B	14	13,486	13,500
		42	27,558	27,600

$$\mathbf{RR_1 = \frac{353.3}{216.7} = 1.6}$$

$$\mathbf{RR_2 = \frac{16.0}{11.7} = 1.4}$$

$$\mathbf{RR_3 = \frac{2.0}{1.0} = 2.0}$$

$$\mathbf{Adj.\,RR = 1.7}$$

Notice that the estimate of the adjusted relative risk (1.7) is the same as that obtained from direct standardization using an external standard, underscoring the correspondence between the two methodologies. With the data organized in this fashion, however, the difference in the stratum specific estimates can also be seen. The process of rate standardization always assumes that a summary measure is appropriate to report, and while it is debatable as to whether effect modification exists in this example, the organization of the data for stratified analysis encourages closer examination of the stratum specific estimates. In this case, it is interesting to see that, the relative risk of neonatal death at the community hospitals compared to the tertiary care centers is actually highest in the normal birthweight category (2.0/1.0=2). While one would expect the community hospitals not to perform as well with high risk infants since they are not equipped to manage them, it is disturbing that their performance is also worse with low risk infants.

Too often, stratum specific information is ignored in favor of the adjusted summary measure. Sometimes this is deemed necessary if, for example, many indicators are being examined, and reporting all of the stratum specific rates may be providing an audience more information than can be easily absorbed. In this circumstance, reporting one summary measure for each indicator may do a better job of communicating a coherent picture of the health status of a population. On the other hand, when the public health focus is on more effective and efficient targeting of interventions and on prioritizing allocation of resources, stratum specific information may in fact be more useful than summary measures.

Moreover, it is important not to confuse the reporting of relative comparisons across populations with examining the true level of an outcome in each population. There is a temptation to try and do both simultaneously by reporting the adjusted rates obtained from the standardization process. A series of such rates implies comparisons across populations, but it also implies that each rate reflects the true level of the outcome in a population which, as was pointed out earlier, it does not. It is better to report an adjusted relative risk or SMR for comparison purposes and in addition report the actual observed stratum specific rates to give a sense of the occurrence of the outcome in each population. In the hospital group example, the adjusted relative risk of 1.7 could be reported along with the stratum specific neonatal mortality rates of 353.3, 16.0, and 2 per 1,000 live births for the community hospitals and then 216.7, 11.7, and 1.0 per 1,000 live births for the tertiary care centers.

Finally, remember that adjustment for confounding is not appropriate when the third factor is in the causal pathway. If the association of interest were smoking status and neonatal mortality instead of hospital group and neonatal mortality, it would be inappropriate to standardize for or stratify by birthweight since birthweight is in the causal path between smoking and neonatal mortality. This illustrates the importance of basing analytic choices on substantive and not mechanical grounds.

❑ Synthetic Estimation

Simple synthetic estimation is part of the process of indirect standardization. Stratum specific rates from either population data, or from a large sample survey are applied to observed numbers in the population of interest. In other words, the denominator of an SMR is a synthetic estimate. This estimate is used when data for an indicator are not collected for the local area and therefore there is no observed value from the local area to use as a numerator for an SMR. A synthetic estimate may also be used when the direct data for the area are available, but known to be under or over-reported, or very unreliable due to small sample size.

Indirect estimates, including SMRs as well as synthetic estimates, are probably biased (in a statistical sense), but they are usually quite reliable since they are derived from very stable rates from large populations or surveys. Recalling that the accuracy of an estimate is dependent on both bias and reliability, using a somewhat biased synthetic estimate may be preferable to using very unstable direct data. And when no direct data are available, the choice is between using a synthetic estimate or none at all.

Suppose we did not have access to direct data for the two hospital groups used in the previous example of standardization, but a comparison of neonatal mortality in the two groups was still desired. If the same hypothetical standard--all live births in the State--were used to calculate separate synthetic estimates for the two hospital groups, the results would be:

$$\text{Synthetic Estimate}_A = 4.39$$

$$\text{Synthetic Estimate}_B = 7.52$$

These are the denominators of the SMRs calculated earlier or the expected rates in the two hospital groups (see pp. 5-6). By using the neonatal mortality experience in the State as a whole as the standard, the assumption is that it is a reasonable reflection of the neonatal mortality experience in each hospital group. Notice, however, that the synthetic estimate for Hospital Group A is lower than that for Hospital Group B, implying that the community hospitals have a better neonatal mortality rate than do the tertiary care centers. This contradicts the results of both direct and indirect standardization as well as the stratified analysis. This kind of result illustrates the danger of using synthetic estimates.

For synthetic estimation to be credible, then, it is critical to use a standard with strata that account for important characteristics of the population for which the estimate is being calculated. With no direct data as a basis of comparison, we must be confident that the stratum specific rates applied to the numbers in the population of interest are close to those that would be observed in that population. To calculate reasonable synthetic estimates for the tertiary care centers and community hospitals, for example, the standard would have to be stratified by hospital type as well as by birthweight—in effect, two distinct standards should be used. The data table for calculating the synthetic estimates would then be organized as follows, with 6 rather than 3 strata for the standard as well as for the hospital groups. Now, the synthetic estimates for the community hospitals can be calculated using only the rates from community hospitals statewide, and the synthetic estimate for the tertiary care centers can be calculated using only the rates from tertiary care centers statewide.

	# of Deaths in the Hospitals of Interest	# of Births in the Hospitals of Interest	Stratum Specific Rates from the Standard	Synthetic Estimates or Expected # of Deaths
Community Hospitals				
< 1500				
1500 – 2499				
>= 2500				
Tertiary Care Centers				
< 1500				
1500 – 2499				
>= 2500				

If an overall synthetic estimate for all hospitals in the region had been desired rather than separate estimates by hospital type, then using all live births in the state as the standard would have been appropriate. Combining the data for Hospital Group A and Hospital Group B into regional totals, we get:

Hospital Groups A and B Combined:

Birthweight	Deaths	Births	%	Stratum Specific Rates from the Standard per 1000	Crude Rate per 1000
< 1500	118	450	1.5	262.0	
1500 – 2499	26	1,950	6.5	13.0	
>= 2500	42	27,600	92.0	1.5	
	186	30,000	100.0		6.2

Remember that the crude rate in each of the hospital groups was 6.2 and therefore the combined crude rate is also 6.2. Multiplying the number of births in the birthweight strata in the above table by the neonatal mortality rates from the hypothetical standard, we get the following synthetic estimate:

$$\mathbf{\frac{450 \times 275 + 1{,}950 \times 14.0 + 27{,}600 \times 1.0}{30{,}000}}$$

$$\mathbf{= 6.0}$$

This synthetic estimate of 6.0 is a reasonable reflection of the actual crude rate in the two hospital groups combined of 6.2.

Following is another example of synthetic estimation. The standard is national survey data stratified on multiple variables to calculate an overall state synthetic estimate of the percent of women who drink alcohol during pregnancy. The strata chosen to calculate the estimate were African American/Non African American, married/not married, and age < 20, 20-34, and >= 35. Separate estimates for African American and Non African American women, or for women in specific age groups could also be calculated analogous to the separate estimates for the community and tertiary care hospitals.

Although direct data have been available on the birth certificate for alcohol use during pregnancy since 1989, it is likely that in the first few years of data collection, the error rate was high. These data are for 1989.

Estimating the Number and Proportion of Women Who Use Alcohol During Pregnancy

Strata	# of Drinkers Reported on the Birth Certificate	# of Live Births	Stratum Specific Percents of Alcohol Use During Pregnancy National Survey Data	Synthetic Estimates or Expected # of Drinkers
African American				
< 20, Married	0	281	0.06	17
< 20, Not Married	99	10,396	0.07	728
20-34, Married	183	7,531	0.11	828
20-34, Not Married	972	18,003	0.16	2,880
> = 35, Married	33	1,002	0.11	110
> = 35, Not Married	77	896	0.13	116
Non African American				
< 20, Married	38	4,264	0.14	597
< 20, Unmarried	134	8,873	0.09	799
20-34, Married	2,761	98,712	0.24	23,691
20-34, Not Married	701	16,083	0.24	3,860
> = 35, Married	584	12,401	0.24	2,976
> = 35, Not Married	67	942	0.25	236
Total	5,649	179,384		36,838

Direct Estimate from Birth Certificate = (5,649/179,384)*100 = 3.1

Synthetic Estimate = (36,838/179,384)*100 = 20.5

In this example, the synthetic estimate of 20.5 % is much higher than the 3.1 % reported in the birth certificate data. We hypothesize that there is underreporting in the birth certificate data, yet we may also not feel confident in the synthetic estimate. Other stratification schemes might be tried to assess any change in the resulting synthetic estimate. In addition, a small local survey may be undertaken in an attempt to get another estimate as a point of comparison.

Should we report either of these less than ideal estimates of alcohol use during pregnancy? Often, it is a matter of public health discretion whether to report estimates in which we do not have full confidence, either because they contain statistical or epidemiologic bias or because they are unreliable. The consequences for public health programs of reporting or not reporting such estimates must be considered. Ironically, there is often more comfort in reporting estimates that are known or suspected to be inaccurate if they are direct estimates, such as poorly reported indicators from vital records, than in reporting indirect (synthetic) estimates that in fact may be more accurate. MCH professionals need to bring their clinical and programmatic knowledge to bear when deciding which indicators are important and which estimates of those indicators to report.